Small-to-Medium-Scale Sensory Evaluation of Horticultural Crops—Measuring Responses[1]

Sean Michael Campbell and Charles A. Sims[2]

Introduction

When measuring the responses of panelists, the main principle behind sensory evaluation, a variety of sensory tests can be used. In an ideal world, sensory panelists would be like machines, capable of being calibrated for maximum precision and accuracy. In reality, all of the factors that contribute to making people unique also cause them to experience sensory attributes differently. The study of these differences is called psychophysics and is a large area of focus for many university psychology departments. While this may seem like an abstract idea, these concepts have effects on our everyday lives. For example, maple syrup, a valuable agricultural commodity, can be rated based on color, clarity, and flavor, with the lighter Grade A varieties selling for a higher market value than the darker Grade B varieties (Figure 1).

This publication is the second in a series designed to assist producers in the small-to-medium-sized sensory evaluation of their horticultural crops, summarizing the types of sensory data available and their associated collection methods, with the guidelines outlined in this publication taken from the 4th edition of *Sensory Evaluation Techniques* (Meilgaard, Civille, and Carr 2016).

Figure 1. Four grades (from left to right, Vermont Fancy, Grade A Medium Amber, Grade A Dark Amber, and Grade B) of maple syrup. Credits: Dvortygirl, Wikimedia Commons

Types of Data

When measuring panelist responses, data can be taken in multiple ways but is commonly collected as nominal, ordinal, interval, or ratio data. Nominal data involves placing items into different, mutually exclusive categories that differ in name but otherwise do not provide a quantitative (numerical) value or follow any order. The top row of Figure 2 shows images of hop (*Humulus lupulus*) cones from three different cultivars classified into three categories, but it otherwise offers little information on the relationships

between them. Nominal data can be a useful tool for differentiating groups but lacks details about how these items relate or differ from each other.

Ordinal data also involves grouping items into distinct categories, but these categories belong to an ordered series. An example might be the degree of browning of hop cones, which can be placed into three categories: light, medium, and heavy. While the ordinal scale and its associated data are limited, they do carry information about relationships between categories.

Interval data is classified by items being placed into distinct categories, separated by a constant interval. In the example given in Figure 2, the three young hop plants are increasing in number of leaves from one to two to three, in each case increasing by an interval of one.

Finally, ratio data involves assigning numbers to indicate how a sample compares to a control (e.g., twice as strong, half as sweet, etc.). As illustrated in Figure 2, the sizes of three hop cones were analyzed; the first cone has a height that is ¾ as tall as the second, which is only ¾ as tall as the third.

Figure 2. Examples for the nominal, ordinal, interval, and ratio scales in hops (*Humulus lupulus*).
Credits: Sean M. Campbell, UF/IFAS

Sensory Measurement Techniques

In addition to the different sensory data types available when performing sensory evaluation, there are also multiple categories for sensory measurement techniques. The most used are classification, grading, ranking, and scaling, with further descriptions listed below in order of increasing technique complexity.

Classification

Classification, the simplest of the sensory measurement techniques, involves placing items being evaluated into groups based on nominal data—that is, differing in name, but not based on numerical values or any sort of order. Most commonly, this involves asking panelists to pick a descriptor from a list that best describes an attribute about the sample in question, often by checking it with an X. Results are reported as the number of checks per response.

Because the data collected is nominal and has no numerical value attached, selecting the proper descriptors is crucial. Most consumers are unfamiliar with the complex descriptors used in sensory evaluation (sucrose, acidity, nondescript), so more common terms must often be used (sweet, sour, bland).

Another problem is a substitution caveat: if panelists find a defect in a product or a noticeable difference between products, and they cannot find an appropriate descriptor for it, they will often select a substitute term on the list to express this, which may not accurately reflect their perceptions. Therefore, it is important to make sure all possible options are provided, which can be done by researching the terms and expressions commonly used by your consumers or by looking up an existing list of appropriate terms created by a trained panel. The following questions in Figure 3 provide examples of nominal data collection:

Classification Example

1. Which attribute best describes the flavor of the strawberry in question:
   _____ sweet   __X__ sour   _____ bland
2. Which attribute best describes the color of the apple in question:
   __X__ red   _____ yellow   _____ green
3. Which attribute best describes the smell of the apple in question:
   _____ fruity   _____ green   __X__ floral

Figure 3. Classification examples using nominal data.
Credits: Sean M. Campbell, UF/IFAS

Grading

Grading involves assigning a value from a known scale to an item based on its performance, just as when assessing classroom assignments. In sensory terms, grading is the placing of items into ordinal data, or distinct groups belonging to an ordered series. Graders use this ordered series to consider all relevant sensory factors about an item before they give one overall rating, usually based off several grade levels, depending on the product. The distinction between grading and ranking is that grading is commonly done by those who learned the craft from others and can now be called expert graders.

Grading is routinely used in meat, dairy, coffee, and tea, and it is primarily a way of protecting the consumer from product substitution or adulteration. Figure 4 illustrates six grades commonly used to describe the degree to which coffee beans have been roasted. By enforcing a series of standards associated with different grades, consumers can be guaranteed that they are getting what they paid for. This also allows producers to earn more money for their products; as an item goes up in grade or as that grade becomes more associated with quality, retailers can charge a premium for that product. There are some drawbacks to the grading system; namely, it requires the use of expert graders trained in the appropriate method, something that takes time and money. For that reason, traditional grading scales are being replaced with more automated sensory evaluation techniques (e.g., mechanization, computer imaging).

Figure 4. Six coffee roasting grades. From top left to bottom right: light cinnamon, cinnamon, normal, French roasting, espresso and open fire.
Credits: Godewind, Wikimedia

Ranking

Like grading, ranking involves placing samples in nominal format, usually based on the intensity or preference of an attribute in question. Subjects are typically given between three to seven samples, and after evaluating them, they are asked to arrange the samples in order based on the intensity of the attribute or preference. If the attribute in question is "Sweetness Intensity," panelists would place the sweetest sample in spot 1, followed by the next sweetest at spot 2, and so on. Rank totals are calculated for each sample.

Ranking tests can be performed rapidly and with very little training, allowing for a wide array of applications. However, ranking tests cannot be used to provide intensity data (Sample A has 30% more intense sweetness than B). The following is an example of ordinal data collection using ranking:

Ranking Test Example

Place the letter of the sample by the corresponding ranking based on sweetness intensity, with 1 being the most intense and 5 the least.

 a. Orange Juice
 b. Apple Juice
 c. Cranberry Juice
 d. Lemonade
 e. Water

 __d__ 1   __b__ 2   __a__ 3   __c__ 4   __e__ 5

Figure 5. Ranking example using ordinal data.
Credits: Sean M. Campbell, UF/IFAS

Scaling

Finally, scaling describes the process of comparing the sample to a predetermined scale, often based on numbers (1–10) or descriptive words (soft-hard). Several scaling method types are available and will yield different types of data, depending on which is used. This publication will focus on two types, line scaling and category scaling. Regardless of the method used, a common issue with scaling is the end-point bias. When using scales, panelists tend to preferentially use points toward the middle of the scale, saving the end points for extreme samples that may never come. Always make sure that your scale is large enough to describe the full range of experiences possible for the crop being tested, while still having enough points on the scale to be able to distinguish minor differences between samples.

Line scaling is the simplest of scaling methods, where panelists are given a line and asked to place an x or other mark on it to match the intensity of the crop or associated attribute. The line has anchors on either end, where commonly the left side stands for a low or zero value and the right end for a high or maximum value (Figure 6).

Line Scaling Example

Place an X on the line to match the intensity of the flavor attributes listed:

1. Sweet — None .... X .... Very
2. Sour — None .... X .... Very
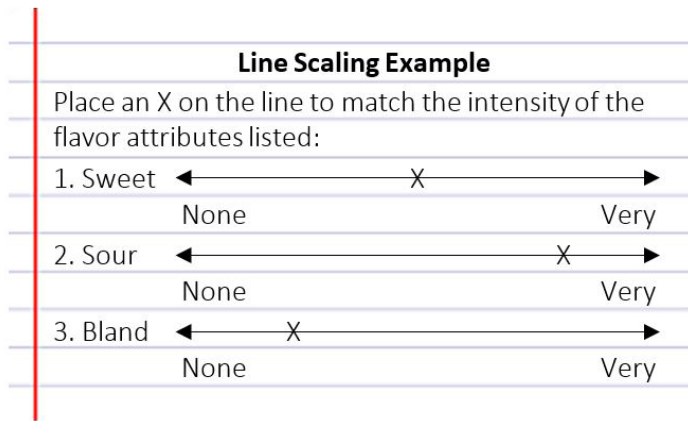3. Bland — None .. X .... Very

Figure 6. Line scaling example using interval data.
Credits: Sean M. Campbell, UF/IFAS

The marks can then be converted to numerical values by measuring their location on the line with a ruler. This type of scaling method is popular because it is very easy for untrained panelists to understand, but without distinct points on the scale for the panelists to reference, user error is often high.

Category scaling is likely one of the more recognizable forms of scaling; it involves asking panelists to rate the intensity of a crop or associated attribute by assigning it a value (category) on a limited scale (Figure 7). This is considered ordinal-level data; however, while the items being tested are placed into distinct groups that belong to an ordered series, it does not provide information on the degree that samples differ by. For example, on a 9-point scale, the difference between a 2 and a 5 rating might not be the same as a difference between a 5 and an 8, or a score of 4 might not indicate a value half as much as a score of 8. Another very popular category scale is the 9-point hedonic scale commonly used to measure overall acceptability of a food. The scale ranges from 1, dislike extremely, to 5, neither like nor dislike, to 9, like extremely.

Summary

To accurately measure consumer perceptions of fruits and vegetables, data can be taken in a variety of forms, increasing in the relative amount of information that can be determined about the relationships between groups. Nominal data involves placing items into different, mutually exclusive categories that differ in name, but otherwise do not provide a quantitative (numerical) value or follow any order. Ordinal data also involves grouping items into distinct categories, but instead these categories belong to an ordered series. Interval data is classified by items being placed into distinct categories, separated by a constant interval. Finally, ratio data involves assigning numbers to

indicate how a sample compares to a control (e.g., twice as strong, half as sweet, etc.).

To collect this sensory data, multiple sensory measurement techniques can be used, each resulting in a different type of data. Classification, the simplest of the sensory measurement techniques, involves placing items being evaluated into groups based on nominal data—that is, differing in name, but not based on numerical values or any sort of order. Grading involves assigning a value from a known scale to an item based on its performance, an example of ordinal data, or distinct groups belonging to an ordered series. Ranking also involves placing samples in nominal format, usually based on the intensity or preference of an attribute in question. Subjects are typically given between three to seven samples, and after evaluating them, they are asked to arrange the samples in order based on the intensity of the attribute or preference. Finally, scaling describes the process of comparing the sample to a predetermined scale, often based on numbers (1–10) or descriptive words (soft-hard).
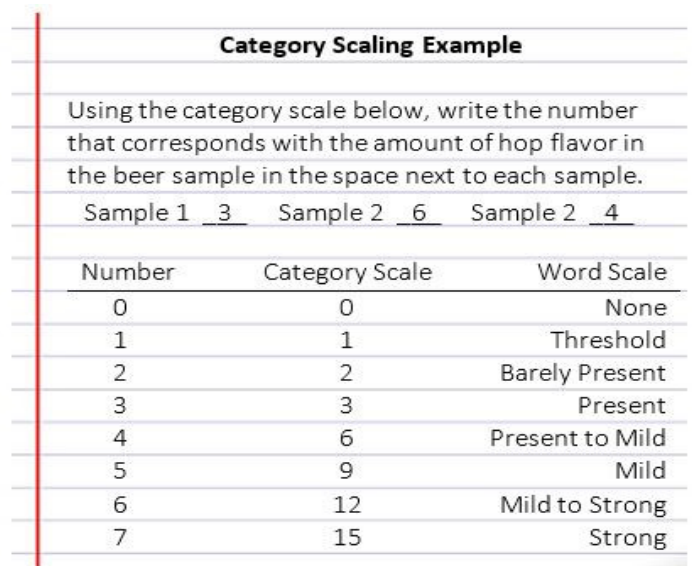
Category Scaling Example

Using the category scale below, write the number that corresponds with the amount of hop flavor in the beer sample in the space next to each sample.

Sample 1 __3__   Sample 2 __6__   Sample 2 __4__

Number	Category Scale	Word Scale
0	0	None
1	1	Threshold
2	2	Barely Present
3	3	Present
4	6	Present to Mild
5	9	Mild
6	12	Mild to Strong
7	15	Strong

Figure 7. Seven-point Likert category scaling example using ordinal and ratio data.
Credits: Sean M. Campbell, UF/IFAS

Reference

Meilgaard, M. C., G. V. Civille, and B. T. Carr. 2016. *Sensory Evaluation Techniques*. Boca Raton: CRC Press.